Combating DDoS Attacks in the Cloud: Requirements, Trends, and Future Directions

Gaurav Somani, Central University of Rajasthan

Manoj Singh Gaur, Malaviya National Institute of Technology

Dheeraj Sanghi, Indian Institute of Technology Kanpur

Mauro Conti, University of Padua

Muttukrishnan Rajarajan, City University London

Rajkumar Buyya, The University of Melbourne

Distributed denial of service (DDoS) attacks targeted to cloud services have serious consequences such as long downtime, economic loss, and both short- and long-term business and reputation losses. We present an overview of these attacks and their variants with respect to cloud infrastructure as well as explain the attack dynamics. Cloud resource management based on autoscaling algorithms is used to build the required DDoS mitigation solutions. These requirements include sustainability or budget constraints, controlled autoscaling, minimization-based optimized control of attack traffic, mitigation throughput time, and service quality and availability. Toward this end, we develop and propose a detailed guideline on possible solutions leading to a novel collaborative solution framework based on multilevel alert flows. We also comment on future attacks in the DDoS space and give a novel DDoS attack variant—detection near impossible DDoS—as an anticipated vision of future attacks to help orchestrate upcoming solutions from the community.

cloud computing, DDoS attacks, security and protection (e.g., firewalls)

DISTRIBUTED DENIAL OF SERVICE (DDoS) ATTACKS HAVE BEEN A NIGHTMARE FOR ENTERPRISE OPERATIONS, AVAILABILITY, AND SECURITY. After the emergence of modern computing paradigms like cloud computing, these attacks saw major changes in scale, methods, aims, and targets. The advantages provided by cloud computing are available to both victims as well as the attackers. This has made the DDoS arms race interesting and quite complex.[1] In 2004, the peak attack bandwidth was just 8 gigabits per second (Gbps). However, according to the report by Arbor Networks, there were much heavier DDoS attacks that had attack bandwidths of more than 500 Gbps in 2015.[2] The target services of DDoS attacks lie in each sector that is influenced by IT infrastructure, whether its government, banking, or media industry.

As per the report by Arbor Networks, the percentage of attacks targeting cloud-based services is growing each year. Over 33% of total reported DDoS attacks in 2015 targeted cloud services, which make the cloud one of the major attack targets. Motivation for the DDoS attacks ranges from extortion, demonstration of attack capabilities, and hacktivism to business rivalry. It is interesting to note the rise of DDoS-attack-for-hire payment-based services, also known as booters or stressers, that attack a target via the planting of attack guns (botnets).[3] With the advent of these methods, the attack frequencies to victim organizations have increased considerably in recent few years. DDoS attacks may last between a few seconds to even weeks in a few cases, which multiplies the economic and business losses multifold.

In particular, the attack duration has enormous impact on the services running on the cloud due to the on-demand utility computing model of the cloud. Financial losses due to DDoS attacks have multiple components or symptoms, few of which are quite visible during the attack. However, the remaining part of the losses are visible only after the attack disappears. Most of these losses are difficult to measure, including the long-term reputation and resultant

business losses. There are recent and much talked about massive DDoS attacks on cloud services and cloud service providers that have shaped the so-called battlefields of the cyberattacks.

The first popular attack was around Christmas 2014 on Sony and Microsoft gaming servers that were used for cloud-based services related to PlayStation and Xbox, respectively. Similarly, there was an attack on a cloud service provider Rackspace that was a DNS DDoS attack and lasted more than 11 h. Another attack in the third quarter of 2014 was a combination of hacking and DDoS attack on Amazon EC2 cloud services.[4] A more recent attack around New Year's eve of 2016 occurred on cloud-hosting provider Linode and lasted for more than a week. These attacks have made the whole cyber security research community think about the scale and strength of these attacks. This introspection and reevaluation of the mitigation methods is also substantiated by the attackers' target shift toward cloud services.

In this paper, we provide a detailed appraisal of the major requirements of efficient DDoS mitigation solutions and the factors governing these requirements. We describe DDoS attack dynamics and cloud resource allocation model to determine the major reasons behind the fatal impact of these attacks on cloud services. We also provide the research gaps and issues associated with the solutions available in the past. This will help in establishing the important requirements for DDoS mitigation solutions and its design considerations.

The paper is organized as follows. The next section details the attack dynamics with a focus of attack methods and consequences on cloud services. We then provide a detailed cloud system analysis of the role of fine- to coarse-grain resources in cloud autoscaling. The following section provides the detailed requirements of cloud-specific mitigation solutions followed by details of the design goals of DDoS mitigation solution, illustrating the proposed collaboration-based multilevel alert flow framework. The final section draws the conclusions of this work.

DDoS Attack Dynamics

Traditionally, analyzing DDoS attacks and designing mitigation solutions have been a hot research area for the cyber security community. Only recently, after the emergence and adoption of cloud computing, DDoS attack research has seen a shift in its focus toward cloud services. Figure 1 details the major attack mechanisms and possible impact of attacks on cloud services and infrastructure. The scale of the DDoS attacks on the cloud is mostly volumetric/massive, having attack bandwidths >100 Gbps. However, there are a few attack incidents of very sophisticated or intelligent attacks where the attackers sent low-rate DDoS attacks to defeat the attack detection mechanisms.[5]

A simple example of such a sophisticated attack is to send attack requests from a large number of sources. The number of attack requests are slightly less than the detection thresholds and so remain undetectable. It is interesting to note that a low-rate DDoS attack with an attack frequency of just one request per minute for a month on a cloud service can be costly.[6] Attack infrastructure used by DDoS attackers for the cloud ranges from Botnets that may range from a few hundred to thousands of malware-affected computers, phones, servers, or cloud virtual machines (VMs) following the instructions of an attacker-governed command-and-control server.

New players in the DDoS attack space are DDoS-for-hire service providers that provide attack as a service with a utility-based business model. The notion of an arms race for DDoS attacks[1] is substantiated by attackers utilizing large amounts of cloud resources. These clouds may be the attacker-administered clouds (termed black or gray clouds) or public clouds hosting attacker VMs. These attacks will make the DDoS mitigation more and more expensive as the resources needed to circumvent these attacks will proportionately grow.

Target protocols/ports may also make the attack effects different. Most of the time, the target services are web services with attack packets of the HTTP GET type. There are other popular attack packets like TCP SYN, ICMP (Internet Control Message Protocol) ECHO, and HTTP POST. Recently, fatal and powerful amplification attacks, employing DNS, NTP (Network Time Protocol), or SNMP (Simple Network Management Protocol), were used that quickly reached high levels of attack. Service disruption remains the primary effect, which is visible in most cases. Economic losses due to the ensuing downtime, disruption to other dependent services/websites, and long- and short-term business reputation losses are the major effects of the DDoS attacks.

In addition to these effects, DDoS attacks in the cloud might not result in service disruption because of flexibility in acquiring further resources as needed. The resulting economic losses might be very heavy as seen in some recent instances.[6–8] Mitigation of a DDoS attack in the cloud also comes at a price as additional resources are required to help in the mitigation; these have associated costs that are proportional to the strength of the attack. In addition to these losses, collateral damage to cohosted VMs, servers, cloud networking, and cloud as a whole might have adverse effects.[9]

These effects are also dependent upon cloud resource allocation algorithms and various related functions such as

VM placement, VM migration (due to resource-freeing attacks), and load balancing. These additional effects range in their performance interference, higher energy consumption needs, and decrease in return on investment. Most of these losses are shown by several important attack characterization studies in the literature, which are summarized in Table 1.

These studies have revealed different kinds of attack effects due to variants of DDoS attacks on cloud services. The major findings of these characterization studies point to the fraudulent consumption of resources as a result of the economic losses incurred by the DDoS attacks. Additional studies such as the one by Somani et al. have shown the effects on DDoS nontargets.[9] These collateral damages are due to the multitenant nature of cloud computing. Resource sharing and resource contention in such environments result in performance issues to cohosted VMs and other cloud resources.

A detailed discussion on attack and threat models for DDoS attacks in cloud computing was presented by Somani et al.[14] Based on the findings, Figure 2 shows a summary of the major contributions related to DDoS attack prevention, detection, mitigation, and recovery.

Cloud Resource Management

A cloud infrastructure may have a number of physical servers (say n) that are ready to host and run VMs. We represent the physical servers as P_i (i = 1 to n). Each physical server P_i has four basic resources: CPU (C), memory (M), disk (D), and network bandwidth (B). These resources are shared among the hosted service instances. These resources also represent the available capacity of a physical server as a resource vector.

$$\text{Capacity}(P_i) = (C_i, M_i, D_i, B_i). \quad (1)$$

An infrastructure cloud runs services in the form of VM instances. A service S_k might have a single VM instance to support the service. If there is an additional resource requirement, the single VM instance service is governed by horizontal scaling (adding/removing resources on the same VM instance on the same server).

In case the required resources are not available on the same physical server, VM instance migration is used to identify a candidate server where the required resources are available. On the other hand, some services might have multiple instances of the same service running behind a load balancer. In this case, additional instances are started on other candidate servers using vertical scaling. Resources allocated to a VM instance I_j are represented as

$$\text{Allotted}(I_j) = (C_j, M_j, D_j, B_j). \quad (2)$$

As a consequence of DDoS attacks, the target service will see heavy usage of the basic resources. This would trigger the cloud autoscaling algorithm to act and get more resources. Acquiring more resources to serve the traffic surge is analogous to capacity planning in performance management problems.

Autoscaling policy see few performance matrices (such as CPU utilization and response time of the service) and act accordingly to respond to the service state. In case of multi-instance service, a general representation of autoscaling policy would be

$$\begin{cases} +I_{\text{add}} & \text{if } \textit{Service.State} = \text{overload} \\ -I_{\text{remove}} & \text{if } \textit{Service.State} = \text{underload} \\ \text{No change} & \text{if } \textit{Service.State} = \text{normal load} \end{cases} \quad (3)$$

Demand and other application performance parameters are the basis of the amount of resources to be added ($-I_{\text{add}}$) or removed ($-I_{\text{remove}}$). The additional requirement of resources for a service S_k will be

$$\text{Requirement}(S_k) = (C_r, M_r, D_r, B_r). \quad (4)$$

This service requirement is met by the available idle resources in the cloud. Idle resources on a physical server are those that are left after allocating resources to the hosted instances. If I_1 and I_2 are two instances, which are already placed on the server P_i, the remaining idle resources on P_i will be

$$\text{Idle}(P_i) = \text{Capacity}(P_i) - \sum_{S=1}^{2} \text{Allotted}(I_s). \quad (5)$$

The requirement of service S_k can be met by a server that requires idle resources. Equation 4 should also be fulfilled for each individual basic resource given in Equation 1.

$$\text{Idle}\left(P_{\text{candidate}}\right) \geq \text{Requirement}(S_k). \quad (6)$$

If the DDoS attack continues even after the addition of another instance, in that case, the newer instances will also face the attack. These instances will also show autoscaling triggers and may result in adding more instances, making the attack effects epidemic.

DDoS Mitigation: Solution Requirements in Cloud Computing

In this section, we highlight the major requirements of a DDoS mitigation solution considering cloud computing infrastructure as a target. Figure 3 shows various factors and their dependency on other important factors while combating DDoS attacks. Governs relationship is very important from the perspective of DDoS solutions for cloud services. Each one of these factors and their associated roles is the basis of the requirements we have identified for DDoS solutions.

Sustainability and Various Costs

Let us assume that the cost of service under attack S_k for attack duration t is $\text{Cost}\left(S_k, \text{attack}\right)$. This includes the cost of downtime, short- and long-term business losses, and all the other costs listed in the previous section on DDoS attack dynamics. The cost of the service would have been only $\text{Cost}\left(S_k, \text{no attack}\right)$ if there was no attack. $\text{Cost}\left(S_k, \text{no attack}\right)$ does not harm the economic sustainability of the VM owner.

The cost of the service while a DDoS attack mitigation system is in place is $\text{Cost}\left(S_k, \text{mitigation}\right)$. This cost includes the cost of the DDoS mitigation system and the cost of additional resources required to perform the mitigation, backup, and recovery. We assume that every organization's budget includes a component for mitigating these attacks considering their sustainability. We take this budgeted item as $\text{Cost}\left(S_k, \text{budget}\right)$. The primary conditions to mitigating DDoS attacks are given in Equations 7 and 8. The cost of attack mitigation should always be lower than the budget/sustainable costs. Additionally, the cost of mitigation should always be lower than the costs while facing the attack or the business gains received if there was no attack.

Requirement 1: Costs

$$\text{Cost}(S_k, \text{budget}) > \text{Cost}(S_k, \text{mitigation}). \quad (7)$$

$$\text{Cost}\left(S_k, \text{mitigation}\right) < \text{Cost}\left(S_k, \text{attack}\right). \quad (8)$$

Autoscaling/Resource Requirements

While DDoS mitigation is happening, one of the most important factors to carefully control is autoscaling policy. The policy of dynamically adding/removing resources may make DDoS attacks in the cloud quite harmful. Typical autoscaling policies (as given in Equation 3) add resources in resource units/VM instances. This should always be done in consonance with the required service quality and costs. Somani et al. have proposed controlled autoscaling to maintain service quality.[15] Dynamic resource allocation-based mitigation has been studied by Yu et al.[16]

The initial resources $\text{Allotted}\left(I_k\right)$ provide quality Q_{good} that would in turn provide the response time R_{good} for N requests (detailed below in the discussion of service quality and availability where Equation 10 shows that the required service quality can only be achieved if the number of requests is matched with the resources). While DDoS occurs, N would become very high and would require large amounts of resource addition to service S_k.

Attack Filtering

If a DDoS mitigation system is in place, it should be able to minimize N that would in turn minimize the additional

resource requirement, which is designated as $\text{Requirement}(S_k)$ in Equation 4. N includes both attack requests A as well as benign requests B.

$$N = A + B. \quad (9)$$

Therefore, the DDoS mitigation mechanism should minimize A. DDoS mitigation at the application layer is mostly done by filtering the good and the bad traffic. This may leave some attackers undetected (false negatives) and some benign requests falsely marked as attack requests (false positives).

Mitigation Throughput Time

Mitigation throughput time (*MTT*) is the time difference between the attack starting time and the time when it is completely mitigated. MTT will decide the cost of all the important factors governing the DDoS attack.

Attack Scale and Duration

Attack scale and duration are two important factors affecting aspects of attack mitigation and resource requirement. Attack duration is controlled by the attackers and may last between a few seconds to hours or days. Attack duration is also dependent on the mitigation methods and their efficiency in detecting and blocking the attacks as early as possible.

Service Quality and Availability

The major reason for a service provider to offer cloud services is availability. Availability is measured in terms of service uptime. Service quality is also a related factor of service health. For web servers, the important service quality criteria is response time or page serve time R. Let us assume that in the capacity-planning phase of the virtualized web server the resources listed in Equation 2 are needed for instance I_j. To make it simple, we represent the service quality by using only the response time R. Response time R_{good} is accepted as representative of good quality of service (Q_{good}) and R_{poor} is accepted as representative of poor service quality (Q_{poor}).

Assuming a single instance service S_k, the response time will be inversely proportional to available resources to the service instance $\text{Allotted}(I_k)$. Additionally, the response time will be higher if there are more requests (N) to be served.

$$R \propto N \quad \text{and} \quad R \propto \frac{1}{\text{Allotted}(I_j)} \quad \text{and} \quad Q \propto \frac{1}{R}. \quad (10)$$

We see that the initial resources given to instance I_k provide Q_{good}. If the service quality deteriorates to Q_{poor}, autoscaling will add more resources (using vertical or horizontal scaling as described by Equation 3). Resource addition will be done per the autoscaling policy to keep the service quality maintained at Q_{good}. We assume that the service quality is directly controlled by the resources allocated to the service seeing the available load.

We also assume that a mechanism exists that can provide the service quality guarantees by using this control. The available load per second is a function of request arrivals each second. The following are the additional requirements of an efficient DDoS mitigation system.

Requirement 2: Optimization Subproblems

We describe the function $\text{Minimize}()$, which is an optimization function minimizing the input fields by controlling various factors governing the field.

$$\text{Minimize}(A). \quad (11)$$

$$\text{Minimize}(R). \quad (12)$$

$$\text{Minimize}(\text{Requirement}(S_k)). \quad (13)$$

$$\text{Minimize}(\text{Cost}(S_k, \text{mitigation})). \quad (14)$$

$$\text{Minimize}(MTT). \quad (15)$$

In addition to the requirements detailed in Equation 7 and Equation 8, Minimize() forms the major requirements of a DDoS mitigation solution in the cloud. Budget/sustainability requirements may be considered as factors while deciding on the maximum budget to be spent on resources, which requires a balance between costs and resources (Equations 13 and 14).

DDoS Attacks in Cloud Services: Lessons Learned and Future Directions

We present the major lessons learned from recent DDoS attacks on cloud services and the connection to the requirements presented in the previous section. Later in this section, we present a multilevel solution hierarchy as well as future attack trends.

1. Traffic filtering: It is well-established that detection based on traffic filtering alone is not sufficient and foolproof. Modern sophisticated attacks evolve by varying their attack features to remain undetected by traffic filters.

2. Managing costs: DDoS attacks are compromising the finances of victim service owners. While designing mitigation solutions, the cost factor is important when managing the sustainability aspects.

3. Services availability: While mitigating DDoS attacks, a mechanism should exist to run services for benign users with minimum or no downtime.

4. Collaboration: Massive volumetric attacks, power attacks, and other sophisticated attacks are not fully detectable at the victim end. There are many other information/alert points in the cloud stack and Internet stack that may help in gaining important information about the likelihood of attacks. These alerts and subsequent actions based on these alerts may prove to be promising to combat attacks.

5. Damage minimization: DDoS mitigation should also provide for minimizing collateral damages as shown in the study by Somani et al.[9] This can be ensured by isolating and monitoring the efforts at other components such as hypervisors and networks.

6. Resource management: DDoS attacks in the cloud has been evaluated as a resource management problem at the victim service end of several studies.[17,18] The major idea behind these solutions is to provide a guarantee of resource contention-free execution of attack mitigation solutions in the presence of the attacks. These resource management-based methods are useful and cost-effective.

Next, we propose a solution hierarchy. They incorporate the detailed requirements in the previous section and the lessons learned and presented above.

Collaboration-Based Multilevel Alert Flow

We propose a collaboration-based multilevel alert flow system, which is shown in Figure 4. We give this as a supporting framework on top of which other efficient mitigation systems can be designed. We identify five important stakeholders in the scenario. These stakeholders include the victim VM, physical server hosting the VM, the network connecting the VM to local and global network, the cloud infrastructure as a whole, and the ISP providing services to the cloud. All these stakeholders are regularly emanating monitored metrics to the Attack Control module, which is a control program running in isolation. The Attack Control module has five major aims related to service health.

Traffic evaluation, monitored autoscaling, service quality, and service availability are performed using the matrices received by five stakeholders. Victim separation and mitigation is done in response to attack detection by the control activity. Once an attack is detected, the mitigation process is leveraged by giving alerts to each stakeholder for necessary action. Additionally, recovery clouds may help in providing needed extra resources and help in isolating a victim server.

The information flow from each stakeholder will help in monitoring the situation and in taking necessary actions. Approaching each individual stakeholder, a victim service may help in collecting history and statistics on application layer traffic. In addition to implementing attack prevention mechanisms such as CAPTCHAs (completely automated public Turing test to tell computers and humans apart), other important alerts about the susceptible attack source may help the Attack Control to take actions at network and ISP levels. Similarly, physical server-based counters can help in getting resource usage patterns, achieving isolation, and controlling horizontal scaling. Network level efforts, including traffic evaluation and filtering at the VM level, server level, and cloud level to ISP level, help in monitoring overall network activity. Filtering mechanisms can be implemented here in addition to the minimal ingress and egress filtering methods.

Network reconfiguration based on Software Defined Network is also important in supporting the implementation of actions by the Attack Control module. Additionally, this level also helps in monitoring VM migrations and identifying their needs. The cloud level may help in providing abstract input on the incident and in controlling autoscaling, power consumption, and the identification of recovery options. Similarly, the ISP level, which connects the outside world, including attackers to the victim network, may help in providing support for top-level decisions and blockage of attacker networks as well as spreading the same to the global ISP consortia. Additionally, this level may also help in anticipating attacks based on trends and in securing backup lines to help in availability.

Future Attacks

We anticipate that the future of DDoS attacks will be from DDoS-for-hire services, with cloud infrastructures and Internet of Things devices becoming major sites for the attack sources as well as targets.[19,20] Volumetric yet sophisticated attacks may utilize knowledge of scaling, resource pricing, application features, and benign users' behavior.

Both parties are considered as armies, where the winning army is mostly the one with more resources. However, we see a different trend with respect to DDoS attacks in cloud computing. Here, the party that spends more on acquiring resources may not be the winning party. On the contrary, the one that instigates the other party to acquire and commit more and more resources may win the game.

This is mostly true where both attackers and victim servers are running on top of cost-driven models. The attacker's aim is to stop the service, and the victim's aim is to keep running the service. The attacker would also like to invest in resources capped with a cost it can support, which is true in the case of DDoS-for-hire services. On the other hand, looking at the cost of each hour spent on a CPU core, victims would also implement prevention methods to make it difficult for attackers to get positive results with the available resources.

Based on traditional volumetric DDoS attacks, we now see a trend where the attacks are becoming sophisticated and are variable rate based. This allows the attackers to remain undetected. Now, we can provide an extreme example of DDoS attack for which detection is impossible by the methods available today, i.e., a detection near impossible (DeNy) DDoS attack. The phrase "near impossible" has been specifically used to imply a hope for detection and direction for the whole cloud security community. After having a look at the large number of attack instances, we see that these forms of attacks are going to appear soon. The launch of such an attack is shown in algorithm *1*.

Algorithm 1: DeNy DDoS

Data: N Attack Sources S_u, $u = 1 to N$

Benign request distribution/pattern B_u for $u = 1 to N$.

Result: Successful attack for duration T

while *Attack!=Successful && Attack Duration<=* T **do**

Learn B_u for N users;

Prepare N Attack sources;

for all the u from 1 to N **do**

S_u follows B_u;

end

end

In this type of attack, attackers send only benign traffic, but from very large number of sources. This results in traffic that is undetectable yet volumetric.[5] We also feel that this attack may be successful even with the number of sources equal to or slightly higher than the maximum parallel connections the target service can support. Figure 5a shows the traffic filter, filtering out attack requests (red is an attack request and green is benign).

On the other hand Figure 5b, shows the DeNy DDoS where the attack remains undetected due to its benignness. The DeNy DDoS follows two important properties:

1. Benignness: The resultant traffic has no anomalies as per the attack detection rules.

2. False alerts: Any detection method that is traffic based would always give false positives even if the method was able to detect any patterns.

DeNy attacks are mostly benign traffic from a very large number of cloud-driven computationally capable sources. They may not be completely similar to today's stealthy attacks, but they are benign requests that if detected by current methods would generate a huge number of false positives/alerts. We anticipate that the future solutions in

mitigating DDoS attacks in the cloud or in general (with respect to DeNy DDoS attacks) will require a thorough re-appraisal and shift in combating DDoS attacks efficiently.

Conclusion

DDoS attacks in the cloud are different from the behavior of attacks on fixed on-premise infrastructure. We provide a detailed introduction to the attack methods, consequences, and attack dynamics. This novel work is an attempt to analyze and gather the important requirements in designing DDoS mitigation solutions for cloud infrastructure. These requirements include optimization of five important factors governing the attack. These factors are sustainability/budget constraints, controlled autoscaling, minimization-based optimization of attack traffic, MTT, and service quality and availability. We also compile important lessons learned pertaining to DDoS attacks in the recent past.

These lessons suggest that traffic filtering alone may not be sufficient to combat DDoS attacks in the cloud environment. We suggest considering sustainability, collaboration, resource management, damage minimization, and availability while handling DDoS attacks in cloud computing. We provide a multilevel alert flow-based collaborative DDoS detection solution framework that may be beneficial in designing efficient mitigation solutions. Finally, based on attack trends, we illustrate future attack surface learning. This is supported by an extreme attack case of DeNy DDoS as an open research problem.

References

1. J. Mirkovic, M. Robinson, and P. Reiher, "Alliance Formation for DDoS Defense," *Proc. 2003 Workshop on New Security Paradigms*, 2003, pp. 11–18.
2. Arbor Networks, "Worldwide Infrastructure Security Report Volume XI," 2015; https://www.arbornetworks.com/images/documents/WISR2016_EN_Web.pdf.
3. J.J. Santanna, R. van Rijswijk-Deij, R. Hofstede, A. Sperotto, M.Wierbosch, L. Zambenedetti Granville, and A. Pras, "Booters: An Analysis of DDoS-as-a-Service Attacks," *2015 IFIP/IEEE Int'l Symp. Integrated Network Management (IM)*, 2015, pp. 243–251.
4. P. Nelson, "Cybercriminals Moving into Cloud Big Time, Report Says," 2015; http://www.networkworld.com/article/2900125/malware-cybercrime/criminals-/moving-into-cloud-big-time-says-report.html.
5. M. Ficco and M. Rak, "Stealthy Denial of Service Strategy in Cloud Computing," *IEEE Transactions Cloud Computing*, vol. 3, no. 1, 2015, pp. 80–94.
6. M.F. Tannian, J. Idziorek, and D. Jacobson, "The Insecurity of Cloud Utility Models," *IT Professional*, vol. 15, no. 2, 2013, pp. 22–27.
7. M. Sides, A. Bremler-Barr, and E. Rosensweig, "Yo-Yo Attack: Vulnerability in Auto-scaling Mechanism," *ACM SIGCOMM Computer Communication Review*, vol. 45, no. 4, 2015, pp. 103–104.
8. B. Saini and G. Somani, "Index Page Based EDoS Attacks in Infrastructure Cloud," *Recent Trends in Computer Networks and Distributed Systems Security*, G.G. Pérez, S.M. Thampi, R. Ko, and L. Shu, eds., Springer Berlin Heidelberg, 2014, pp. 382–395.
9. G. Somani, M.S. Gaur, D. Sanghi, and M. Conti, "DDoS Attacks in Cloud Computing: Collateral Damage to Non-targets," *Computer Networks*, vol. 109, no. 2, 2016, pp. 157–171.
10. R.M. Mohammad, C. Mauro, and L. Ville, "EyeCloud: A BotCloud Detection System," *Proc. 5th IEEE Int'l Symp. Trust and Security in Cloud Computing (IEEE TSCloud 2015)*, Helsinki, Finland, 2015.
11. J. Latanicki, P. Massonet, S. Naqvi, B. Rochwerger, and M. Villari, "Scalable Cloud Defenses for Detection, Analysis and Mitigation of DDoS Attacks," *Towards the Future Internet*, IOS Press, 2010, pp. 127–137.
12. Z. Xu, H.Wang, Z. Xu, and X.Wang, "Power Attack: An Increasing Threat to Data Centers," *Proc. Network and Distributed System Security Symp. (NDSS)*, vol. 14, 2014.

13. H. Liu, "A New Form of DOS Attack in a Cloud and Its Avoidance Mechanism," *Proc. 2010 Workshop on Cloud Computing Security*, 2010, pp. 65–76.

14. G. Somani, M.S. Gaur, D. Sanghi, M. Conti, and R. Buyya, "DDoS Attacks in Cloud Computing: Issues, Taxonomy, and Future Directions," e-print arXiv: 1512.08187, 2015.

15. G. Somani, A. Johri, M. Taneja, U. Pyne, M.S. Gaur, and D. Sanghi, "DARAC: DDoS Mitigation Using DDoS Aware Resource Allocation in Cloud," *ICISS 2015 Proc. 11th Int'l Conf. Information Systems Security*, vol. 9478, 2015, pp. 263–282.

16. S. Yu, Y. Tian, S. Guo, and D.O. Wu, "Can We Beat DDoS Attacks in Clouds?" *IEEE Trans. Parallel and Distributed Systems*, vol. 25, no. 9, 2014, pp. 2245–2254.

17. G. Somani, M. S. Gaur, D. Sanghi, M. Conti, and R. Buyya, "Service Resizing for Quick DDoS Mitigation in Cloud Computing Environment," *Ann. Telecommunications*, 2016, pp. 1–16.

18. G. Somani, M.S. Gaur, D. Sanghi, M. Conti, and M. Rajarajan, "DDoS Victim Service Containment to Minimize the Internal Collateral Damages in Cloud Computing," *Computers & Electrical Engineering*, 2016.

19. M. Krell, "DDoS and 3 Recommendations to Secure the Internet of Things (IoT)," 2016; http://www.forbes.com/sites/moorinsights/2016/10/26/ddos-and-3-recommendations-to-secure-the-internet-of-things-iot/n#16ae1c022c16.

20. M. Karami, Y. Park, and D. McCoy, "Stress Testing the Booters: Understanding and Undermining the Business of DDoS Services," *Proc. 25th Int'l Conf. on World Wide Web (WWW)*, 2016, pp. 1033–1043.

GAURAV SOMANI *(gaurav@curaj.ac.in) is an assistant professor at department of computer science and engineering at the Central University of Rajasthan, India. He has completed his MTech in information and communication technology from DAIICT, Gandhinagar, India, and BE in information technology from the University of Rajasthan, India, both with a distinction. He is pursuing his PhD from MNIT, Jaipur, India. His research interests include distributed systems and security engineering. He has published several papers in various international conferences and journals and is a reviewer of many top journals. He is also part of program committees of various conferences and was a keynote and tutorial chair for ICISS 2016. He is a member of IEEE and ACM.*

MANOJ SINGH GAUR *(gaurms@gmail.com) is a professor in the department of computer science and engineering at Malaviya National Institute of Technology Jaipur, India. He has obtained his PhD from University of Southampton, UK. He has supervised research in the areas of networks on chip and information security. He has published over 150 papers in peer-reviewed major conferences and journals and has coordinated national and international projects in the domains of information security and networks on chip. He has been associate editors with CSI Transaction, IET Electronics and Digital Techniques, and Journal of Information Security and Assurance. He was organizing chair of SPACE 2015 and was general cochair of SINCONF 2012 and ICISS 2016. He is a member of IEEE and ACM.*

DHEERAJ SANGHI *(dheeraj@iitk.ac.in) is a professor of computer science and engineering at IIT Kanpur. Professor Sanghi has a BTech from IIT Kanpur and MS and PhD from University of Maryland, US. Since August 15, he has started working with IIIT Delhi. From 2008 to 2010, he served as the director of LNM Institute of Information Technology (LNMIIT), a public-private partnership university in Jaipur. His research interests include network performance optimization, security and distributed systems. He has published a large number of papers in major international conferences and journals. He was a general chair at ICISS 2016 and a program chair at IC3 2016.*

MAURO CONTI *(conti@math.unipd.it) is an associate professor at the University of Padua, Italy. He obtained his PhD from Sapienza University of Rome, Italy, in 2009. After his PhD, he was a postdoctoral researcher at Vrije Universiteit Amsterdam, The Netherlands. In 2011, he joined the University of Padua as assistant professor, and in 2015, he became associate professor. He has been visiting researcher at GMU (2008, 2016), UCLA (2010), UCI (2012, 2013, 2014), TU Darmstadt (2013), UF (2016), and FIU (2015, 2016). He has been awarded with a*

Marie Curie Fellowship (2012) by the European Commission, and with a fellowship by the German DAAD (2013). His main research interest is in the area of security and privacy. In this area, he published more than 150 papers in top international peer-reviewed journals and conferences. He is associate editor for several journals, including IEEE Communications Surveys & Tutorials and IEEE Transactions on Information Forensics and Security. He was program chair for TRUST 2015, ICISS 2016, and WiSec 2017, and general chair for SecureComm 2012 and ACM SACMAT 2013. He is Senior Member of IEEE.

MUTTUKRISHNAN RAJARAJAN *(r.muttukrishnan@city.ac.uk) received his BEng and PhD degrees from City University London in 1994 and 1999, respectively. From 1999, he worked as a research fellow at City University London. In August 2000, he moved to Logica as a telecommunication consultant. After a few years in the industry, Raj is now a professor of security engineering. He is a senior member of IEEE, a member of IET, and an associate member of the Institute of Information Security Professionals (IISP) and a member of Technical Programme Committees for PIERS 2010, eHealth 2010, SECURECOM2011, TrustBus 2011, Digital Economy 2012, IFIPTM 2012, and IFIP SEC 2012. He is also the general chair of SECURECOMM 2011 in London. He also sits on the editorial boards of Springer/ACM Journal on Wireless Networks, Elsevier Journal of Health Policy & Technology, and Emerald Journal of Information Management & Computer Security.*

RAJKUMAR BUYYA *(rbuyya@unimelb.edu.au) is a Fellow of IEEE, professor of computer science and software engineering, future fellow of the Australian Research Council, and director of the Cloud Computing and Distributed Systems (CLOUDS) Laboratory at the University of Melbourne, Australia. He has authored over 500 publications and four text books including Mastering Cloud Computing published by McGraw Hill, China Machine Press, and Elsevier/Morgan Kaufmann for Indian, Chinese, and international markets, respectively. He is one of the most highly cited authors in computer science and software engineering worldwide (h-index=105, g-index=221, and 53,100-plus citations). A scientometric analysis of cloud computing literature by German scientists ranked Dr. Buyya as the world's top-cited author and the world's most-productive author in cloud computing. Software technologies for grid and cloud computing developed under Dr.Buyya's leadership have gained rapid acceptance and are in use at several academic institutions and commercial enterprises in 40 countries around the world. He served as the founding editor-in-chief of IEEE Transactions on Cloud Computing. He is currently serving as co-editor-in-chief of Journal of Software: Practice and Experience, which was established 40-plus years ago.*

FIGURE 1. DDoS attacks and losses in cloud services.

FIGURE 2. Various methods of combating DDoS attacks in cloud computing.

FIGURE 3. Who governs what while combating DDoS attacks?

FIGURE 4. DDoS attacks in the cloud: multilevel solution.

FIGURE 5. Traffic filters and undetectable attacks. (a) Traffic filtering. (b) DeNy attacks passing through the filters.

TABLE 1. DDoS attack variants in the cloud.

DDoS Attack Variants	Attack Specialties
Economic denial of sustainability and fraudulent resource consumption6	On-demand resource scaling and economic loss due to fake resource surge
Yo-Yo attack7	Exploits cloud autoscaling vulnerabilities

Stealthy/energy DDoS attacks5	Sophisticated attacks targeting energy consumption losses
Internal DDoS attacks/BotCloud10,11	Cloud-originated attacks/cloud-originated attacks
Collateral damage to nontargets9	Performance impacts on nontargets in multitenant clouds
Power meltdown12	Energy consumption
Index page EDoS* attack8	Attacks on index pages
Bandwidth DDoS attack13	Bandwidth-depletion attacks

*EDoS, economic denial of sustainability